National Exams May 2019

17-Ind-A6, Systems Simulation

3 hours duration

NOTES:

1. If doubt exists as to the interpretation of any question, the candidate is urged to submit with the answer paper, a clear statement of any assumptions made.

2. This is a closed book exam. Candidates are permitted to use one of the two permitted calculators (Sharp or Casio models).

3. Candidates are permitted to have an aid sheet consisting of one 8.5" x 11.0" sheet of paper. Writing is permitted on both sides of the paper.

4. This exam consists of three sections (A, B, and C). Within each section, candidates will be given a choice of questions to answer. Please read the instructions for each section carefully. A breakdown of questions and marks is as follows:

Section A:	Do 3 of 5 Questions.	Total marks: 30
Section B:	Do 1 of 2 Questions.	Total marks: 20
Section C:	Do 1 of 2 Questions.	Total marks: 20
Exam:	5 Questions.	Total marks: 70

4. The value of each question is listed in the exam. Remember to check the instructions for each section. DO NOT ATTEMPT TO DO ALL QUESTIONS.

5. Statistical tables are provided.

Part A: Concepts

Complete **three** of the following **five** sets of questions.
Do NOT attempt all questions.
Please note that all questions have the same value.

1. Suppose that X is a continuous random variable with probability density function:

$$f(x) = x^2 + 2/3x + 1/3 \qquad \text{for } 0 <= x <= c$$

a) Prove that c must be 1 for f(x) to be a valid probability function.
b) Calculate E(X) and Var(X)

10 Marks

2. A linear congruential linear random number generator has the following parameters:

$$x_0 = 12$$
$$m = 16$$
$$c = 13$$
$$a = 13$$

a. Determine the next two random numbers that will be returned by this generator
b. What is the maximum possible period of this generator? Why?
c. What do we mean by "repeatability" when we refer to a random number generator? Why is having a repeatable generator important for a simulation model? How is a repeatable random number stream assured in a simulation?

10 Marks

3. Roxanne Potvin is the manager of engineering for New Sluice Power, Inc. Ms. Potvin's group is responsible for providing maintenance services on power lines. Before the start of each project, a team engineer estimates the time required to complete the work (in person months). At the end of the project, the actual time to complete the project is recorded. Assume that Ms. Potvin is interested in modelling the accuracy of the project estimation activity. Accordingly, she takes a sample of 250 projects at random and measures the difference between actual completion time and estimated completion time. Some of the summary statistics for this sample appear below.

(Actual - Estimated)	
Mean	1.001
Sample Variance	0.950
Median	0.991
Skewness	0.004
Range	5.136
Minimum	-5.081
Maximum	6.860
Count	250

a. Even though you don't have access to the original data, suggest a potential distribution for the data summarized above and provide your rationale for suggesting it. There are two or three clues in the data that should give you an idea of its potential distribution.
b. Calculate a 95% confidence interval for the difference between projected and actual project completion.
c. If a project completed 3.0 person months ahead of its projected completion date, would that project be considered an anomaly? Why or why not?

10 Marks

4. Consider the following function:

$$f(x) = \begin{cases} \frac{2x}{9} & 0 \le x \le 3 \\ 0 & Otherwise \end{cases}$$

a) Prove that f(x) is a valid probability function.
b) Develop an inverse-transformation for this function.
c) Assume a multiplicative congruential random number generator with parameters: a = 23, m = 100, and $x_0 = 99$. Generate two random variates from the function for F(x).

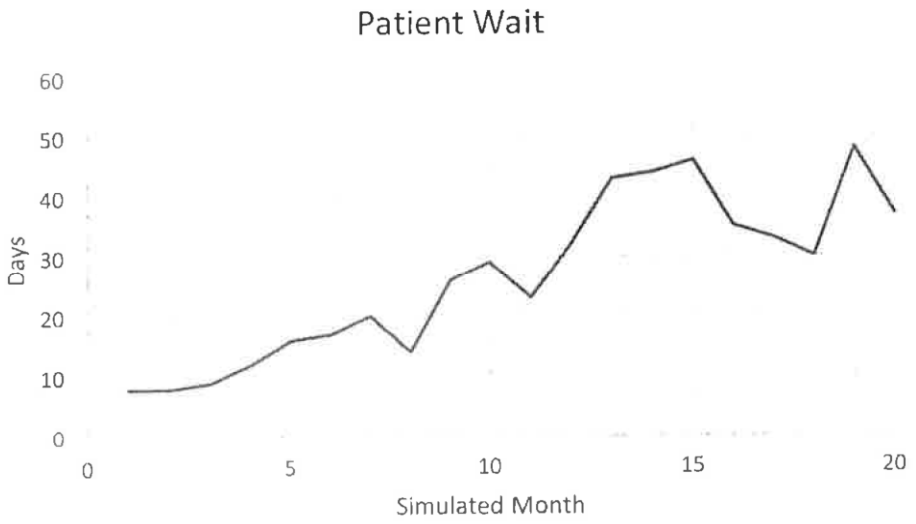
5. Answer the following questions about distributions:
   a. A normal distribution models _____ and _____
   b. The sum of some number of exponential distributions is modelled by _____
   c. If data is thought to be lognormally distributed, parameters can be fit by _____ and assuming the data to be _____
   d. Inverse transformations can be found for all random distributions is a (T/F) _____ statement.
   e. If the number of arrivals per hour follows a Poisson distribution, then the inter-arrival time is _____ distributed.
   f. Poisson distributions model _____ arrivals from a _____ population.
   g. The exponential distribution is a special case of the _____ distribution and the _____ distribution.
   h. The number of items inspected before finding a defective unit is modelled by the _____ distribution.
   i. I draw batches of n items from a large population. The number of defective items in the batch, given a known defective rate (p) is modelled by the _____ distribution.
   j. A triangular distributions is used to model _____.

10 Marks

Part B: Methods

Complete **two** of the following **three** sets of questions.
Do NOT attempt all questions.
Please note that all questions have the same value.

1. Answer the following questions about run length for a simulation model evaluating patient wait to see a physician in a local emergency department.
   a. Why is it important to ensure that a simulation model is run for a sufficiently long time?
   b. Consider two methods of running a simulation model: batch means and deletion/replication. Define both and discuss the importance of setting appropriate run length in the context of each method.
   c. Describe an algorithm to identify the appropriate length of a simulation run under both sets of run assumptions (i.e. under batch means and replication/deletion)
   d. Assume that the output from a simulation run is as follows:



Why would it be difficult, given this graph, to determine an appropriate warm-up period? I'm not looking for calculations here, but a reasoned statement.

   e. Now assume that because I can't find an appropriate run length, I decide to abandon the idea of batch means and employ deletion/replication using an arbitrary run length of 12 months. Is this a good compromise, given a difficult problem? Why or why not?

15 Marks

2. An IE has finished building a simulation model of a hospital emergency department. The IE now wishes to prove that her model is a valid representation of the actual operations of the hospital in question. She has collected data on the average daily number of admissions since 2000:

Actual System	
Year	**Avg Daily Admissions**
2000/01	73.39
2001/02	69.98
2002/03	71.63
2003/04	71.67
2004/05	72.28
2005/06	72.36
2006/07	70.29

The IE wishes to compare this data to a sample of data taken from her simulation:

Simulation Response	
Replication	**Avg Daily Admissions**
1	70.76
2	70.19
3	71.39
4	70.07
5	71.01
6	70.33
7	70.27

a. Is the variance of the simulation statistically different from that of the actual system?
b. If the IE uses $\alpha = 0.05$, can she conclude that her simulation model is an accurate representation of the operations of the hospital?
c. What cautions (if any) do you have for the IE with respect to your conclusion about validity?
d. If you compare a simulation to an actual system and the variances are statistically different, but the means were not, could you say that your model is valid? Why or why not?

15 Marks

3. Erica Strange, an IE from NS Flowers, Inc (NSF) has found, in test runs of a recently completed simulation model, that computation time to complete a run is quite long. She is interested in implementing a variance reduction technique to reduce computation time.

   a. Describe, in your own words, how variance reduction techniques can be used to reduce computation time.
   b. Describe how "common random numbers" can be used to reduce variance.
   c. Describe how "antithetic random variates" can be used to reduce variance.
   d. What are the potential drawbacks (or difficulties) with introducing variance reduction techniques into a simulation project?
   e. Under what conditions can variance reduction techniques involving either common-random numbers or antithetic random variates be assured to work? Are such conditions realistic for all models?

15 Marks

Part C: Applications

Complete **one** of the following **two** sets of questions.
Do NOT attempt all questions.
Please note that all questions have the same value.

1. Assume that a simulation model of an industrial process has just been completed. The simulation is run under three different operating scenarios and the results appear below:

Observation	Scenario A	Scenario B	Scenario C	
1	96	104	141	
2	109	87	169	
3	102	122	77	
4	97	85	87	
5	92	115	94	
Sum	*496*	*513*	*568*	*1577*
Mean	*99*	*103*	*114*	*105*
St.Dev	*6.53*	*16.47*	*39.53*	*24.01*

a) Is there any statistical evidence that the scenarios influence the output metric?
b) Outline all assumptions that underlie your analysis in (a).
c) Which, of any, of the scenarios is best, assuming that bigger is better for the particular output metric of interest? You may pick an appropriate certainty level of your choice.
d) There are at least two methods for determining the "best of k" systems. Define the method that you **did not use in part a** and describe (in words) how you would apply this method to the above data. Do not give extra calculations, just describe how you would complete the calculations.

20 Marks

2. An IE is interested in setting up a screening experiment to identify significant (and insignificant) factors affecting throughput in an industrial process. Three factors have been identified:

- Factor 1: WIP levels (5, 10)
- Factor 2: Queue discipline (FIFO, LIFO)
- Factor 3: Arrival process (Scheduled vs. Random)

a. Design a full factorial (2^k) set of experiments that will test for main factor effects and interaction effects. In your solution, provide a *brief* description of the process you would follow to complete the set of experiments.

b. Assume that the IE has completed five replications of the eight runs laid out in the factorial design matrix in (a). The following results are obtained:

Run	Rep 1	Rep 2	Rep 3	Rep 4	Rep 5	Sum	Avg	Var
1	14	16	11	15	18	74	14.8	6.7
2	21	14	12	14	13	74	14.8	12.7
3	14	16	17	18	15	80	16	2.5
4	17	14	14	18	15	78	15.6	3.3
5	13	14	14	14	12	67	13.4	0.8
6	14	12	12	12	16	66	13.2	3.2
7	11	11	11	14	12	59	11.8	1.7
8	14	17	14	15	17	77	15.4	2.3

Determine which of three factors are significant, using an Analysis of Variance (ANOVA) model and an α value of 0.05. Can you say anything about interaction between any of the factors? To assist in your calculations, you may assume that the Grand Sum of overall runs and replications is 575 and that the variance of the entire sample is 5.266.

c. Describe, in your own words, how the results of the designed experiment will impact the search for a good solution that minimizes transit time?

20 Marks